Long-Run Accuracy of Variational Integrators in the Stochastic Context

Nawaf Bou-Rabee[*]   Houman Owhadi[†]

January 13, 2010

Abstract

This paper presents a Lie-Trotter splitting for inertial Langevin equations (Geometric Langevin Algorithm) and analyzes its long-time statistical properties. The splitting is defined as a composition of a variational integrator with an Ornstein-Uhlenbeck flow. Assuming the exact solution and the splitting are geometrically ergodic, the paper proves the discrete invariant measure of the splitting approximates the invariant measure of inertial Langevin to within the accuracy of the variational integrator in representing the Hamiltonian. In particular, if the variational integrator admits no energy error, then the method samples the invariant measure of inertial Langevin without error. Numerical validation is provided using explicit variational integrators with first, second, and fourth order accuracy.

[*]Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012-1185 (`nawaf@cims.nyu.edu`).

[†]Applied & Computational Mathematics and Control & Dynamical Systems, Caltech, Pasadena, CA 91125 (`owhadi@acm.caltech.edu`).

1   Introduction

Overview   This paper analyzes equilibrium statistical accuracy of discretizations of inertial Langevin equations based on variational integrators. Variational integrators are time-integrators adapted to the structure of mechanical systems [9]. The theory of variational integrators includes discrete analogs of the Lagrangian, Noether's theorem, the Euler-Lagrange equations, and the Legendre transform. Variational integrators can incorporate holonomic constraints (via, e.g., Lagrange multipliers) [26] and multiple time steps to obtain so-called asynchronous variational integrators [8].

The generalization of variational integrators the paper analyzes are derived from a Lie-Trotter splitting of inertial Langevin equations into Hamiltonian and Ornstein-Uhlenbeck equations. The integrator is then defined by selecting a variational integrator to approximate the Hamiltonian flow and using the exact Ornstein-Uhlenbeck flow. Such a generalization of variational integrators to inertial Langevin equations will be called a Geometric Langevin Algorithm (GLA).

This type of splitting of inertial Langevin equations is natural, but seems to have been only recently introduced in the literature (for molecular dynamics see [7, 25], for dissipative particle dynamics see [18, 19], and for inertial particles see [16]). This paper is geared towards applications in molecular dynamics where inertial Langevin integrators (including the ones cited above) have been based on generalizations of the widely used Störmer-Verlet integrator. The Störmer-Verlet integrator is attractive for molecular dynamics because it is an explicit, symmetric, second-order accurate, variational integrator for Hamilton's equations. In molecular dynamics it was popularized by Loup Verlet in 1967. Other popular generalizations of the Störmer-Verlet integrator to inertial Langevin equations include Brünger-Brooks-Karplus (BBK) [5], van Gunsteren and Berendsen (vGB) [24], and the Langevin-Impulse (LI) methods [20]. The LI method is also based on a splitting of inertial Langevin equations, but it is different from the splitting considered here. To our knowledge there are few results in the literature which quantify the long-time statistical accuracy of the Lie-Trotter splitting considered here.

GLA is not only quasi-symplectic as defined in RL1 and RL2 of [14], but also conformally symplectic, i.e., preserves the precise symplectic area change associated to the flow of inertial Langevin processes [12]. One way to prove this property is by deriving the scheme from a variational principle and analyzing its boundary terms as done in the context of stochastic Hamiltonian

2

systems without dissipation in [3].

Organization of the Paper   In §2 the main results of the paper are presented. §3 states all of the hypotheses used in the paper. These hypotheses are invoked in §4 where it is proved that GLA is pathwise convergent on finite time intervals (Theorem 2.1), GLA is geometrically ergodic with respect to a nearby invariant measure on infinite time intervals (Theorem 2.2), and the equilibrium statistical accuracy of GLA is governed by the order of accuracy of the variational integrator in representing the Hamiltonian (Theorem 2.3). In §5, numerical validation is provided. In the Appendix we review some basic facts on variational integrators for the reader's convenience.

Limitations   In a nutshell the main result of the paper states that if GLA is geometrically ergodic with respect to a unique invariant measure, the error in sampling the invariant measure of the SDE is determined by the energy error in GLA's variational integrator. Now if the inertial Langevin equations have nonglobally Lipschitz drift and the GLA is based on an explicit variational integrator, GLA may fail to be geometrically ergodic. In particular, for any step-size there will be regions in phase space where the Lipschitz constant of the drift is beyond the linear stability threshold of GLA's underlying variational integrator. Hence, an explicit GLA will be stochastically unstable. Since our results rely on a strong form of stochastic stability of GLA (namely, geometric ergodicity), they may not hold in this case.

To stochastically stabilize GLA, one can use GLA as a proposal move in a Metropolis-Hasting method. For a numerical analysis of the Metropolis-adjusted scheme, the reader is referred to [4]. A difficulty in Metropolizing inertial Langevin is that its solution is not reversible. However, the solution composed with a momentum flip is reversible. The role of momentum flips in Metropolizing Langevin integrators is qualitatively and computationally analyzed in [1, 7, 17]. For a quantitative treatment of the role of momentum flips in pathwise accuracy the reader is referred to [4].

Extension to manifolds   For the sake of clarity, the setting of this paper is inertial Langevin equations on a flat space, but we stress GLA and its properties generalize to manifolds. We refer to Remark 2.1 and to [2] for details.

3

2   Main Results of Paper

Inertial Langevin The setting of the paper is a dissipative stochastic Hamiltonian system (as in [21, 23]) on $\mathbb{R}^n$, with phase space $\mathbb{R}^{2n}$, and smooth Hamilton $H \in C^\infty(\mathbb{R}^{2n}, \mathbb{R})$. In terms of which consider the following inertial Langevin equations

$$
\begin{cases}
d\mathbf{Y} & = \mathbb{J}\nabla H(\mathbf{Y})dt - \gamma \boldsymbol{C}\nabla H(\mathbf{Y})dt + \sqrt{2\gamma\beta^{-1}}\boldsymbol{C}d\mathbf{W} \\
\boldsymbol{Y}(0) & = \boldsymbol{x} \in \mathbb{R}^{2n}
\end{cases}
\tag{1}
$$

where the following matrices have been introduced:

$$
\mathbb{J} = \begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I} \end{bmatrix}.
$$

Here $\boldsymbol{W}$ is a standard $2n$-dimensional Wiener process, or Brownian motion, $\beta > 0$ is a parameter referred to as the inverse temperature, and $\gamma > 0$ is referred to as the friction factor. We will often write the continuous solution in component form as $\mathbf{Y}(t) = (\boldsymbol{Q}(t), \boldsymbol{P}(t))$ where $\boldsymbol{Q}(t)$ and $\boldsymbol{P}(t)$ represent the instantaneous configuration and momentum of the system, respectively. We shall assume the Hamiltonian is separable and quadratic in momentum:

$$
H(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T \boldsymbol{M}^{-1}\boldsymbol{p} + U(\boldsymbol{q}),
$$

where $\boldsymbol{M}$ is a symmetric positive definite mass matrix and U is a potential energy function. Despite the degenerate diffusion in (1), under certain regularity conditions on U, the solution to this SDE is geometrically ergodic with respect to an invariant probability measure μ with the following density [23]:

$$
\pi(\boldsymbol{q}, \boldsymbol{p}) = Z^{-1}\exp\left(-\beta H(\boldsymbol{q}, \boldsymbol{p})\right),
\tag{2}
$$

where $Z = \int_{\mathbb{R}^{2n}} \exp\left(-\beta H(\boldsymbol{q}, \boldsymbol{p})\right) d\boldsymbol{q}d\boldsymbol{p}$. The invariant measure μ is known as the *Boltzmann-Gibbs measure*.

Geometric Langevin Algorithm   Let N and h be given, set $T = Nh$ and $t_k = hk$ for $k = 0, ..., N$. Observe that the conservative part of (1) defines Hamilton's equations for the Hamiltonian H:

$$d\boldsymbol{Y} = \mathbb{J}\nabla H(\boldsymbol{Y})dt$$

or,

$$\begin{cases} d\boldsymbol{Q} & = \boldsymbol{M}^{-1}\boldsymbol{P}dt \\ d\boldsymbol{P} & = -\nabla U(\boldsymbol{Q})dt \end{cases} \qquad (3)$$

Let h be a fixed step-size. We apply a *pth*-order accurate variational integrator, $\theta_h : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, to approximate the Hamiltonian flow of (3) ($p \geq 1$). The nonconservative part of the inertial Langevin equation defines an Ornstein-Uhlenbeck process in momentum governed by the following linear SDE:

$$d\mathbf{Y} = -\gamma\boldsymbol{C}\nabla H(\mathbf{Y})dt + \sqrt{2\gamma\beta^{-1}}\boldsymbol{C}d\mathbf{W}$$

or,

$$\begin{cases} d\boldsymbol{Q} & = 0 \\ d\boldsymbol{P} & = -\gamma\boldsymbol{M}^{-1}\boldsymbol{P}dt + \sqrt{2\beta^{-1}\gamma}d\boldsymbol{W} \end{cases} \qquad (4)$$

Reference [16] aptly refers to (4) as a Gaussian SDE since its stationary distribution on $\mathbb{R}^{2n}$ is Gaussian in momentum.

The following stochastic evolution map $\psi_{t_k+h,t_k} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ defines the stochastic flow of (4):

$\psi_{t_k+h,t_k} :$
$$(\boldsymbol{q},\boldsymbol{p}) \mapsto \left(\boldsymbol{q}, e^{-\gamma\boldsymbol{M}^{-1}h}\boldsymbol{p} + \sqrt{2\beta^{-1}\gamma}\int_{t_k}^{t_k+h} e^{-\gamma\boldsymbol{M}^{-1}(t_k+h-s)}d\boldsymbol{W}(s)\right), \quad (5)$$

with $\psi_{s,s}(\boldsymbol{x}) = \boldsymbol{x}$ and for $0 \leq r \leq s \leq t$ recall the Chapman-Kolmogorov identity $\psi_{t,s} \circ \psi_{s,r}(\boldsymbol{x}) = \psi_{t,r}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^{2n}$. For the distribution of the solution, the stochastic flow will be denoted simply by ψ_h. To make this map explicit, let $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and set

$$\boldsymbol{\Sigma}_h := 2\beta^{-1}\gamma\mathbb{E}\left\{\left(\int_0^h e^{-\gamma\boldsymbol{M}^{-1}(h-s)}d\boldsymbol{W}(s)\right)\left(\int_0^h e^{-\gamma\boldsymbol{M}^{-1}(h-s)}d\boldsymbol{W}(s)\right)^T\right\}$$
$$= \beta^{-1}\left(\boldsymbol{I} - \exp(-2\gamma\boldsymbol{M}^{-1}h)\right)\boldsymbol{M}$$

and define A_h to be the decomposition matrix arising from the Cholesky factorization of $\boldsymbol{\Sigma}_h$, i.e., $\boldsymbol{A}_h \boldsymbol{A}_h^T = \boldsymbol{\Sigma}_h$. In terms of these, introduce the following flow map:

$$\psi_h : (\boldsymbol{q}, \boldsymbol{p}) \mapsto \left(\boldsymbol{q}, e^{-\gamma \boldsymbol{M}^{-1} h} \boldsymbol{p} + A_h \boldsymbol{\xi}\right). \tag{6}$$

In distribution (6) is identical to (5).

Given $\boldsymbol{X}_k \in \mathbb{R}^{2n}$ and h, the *Geometric Langevin Algorithm* (GLA) is defined as the following Lie-Trotter splitting integrator for (1):

$$\boldsymbol{X}_{k+1} := \theta_h \circ \psi_{t_k+h, t_k}(\boldsymbol{X}_k) \tag{7}$$

for $k = 0, ..., N-1$ with $\boldsymbol{X}_0 = \boldsymbol{x}$.

Remark 2.1. *Observe that GLA generalizes to inertial Langevin equations on a manifold. This generalization is possible because its symplectic component can be defined as a variational integrator for Hamilton's equations on a manifold and its Ornstein-Uhlenbeck component can be defined as the solution of an SDE on a vector space. This generalization is motivated by molecular systems with holonomic constraints. As mentioned in the introduction, variational integrators can incorporate holonomic constraints. In the special case that the configuration manifold of GLA is compact (e.g., $SO(3)$) and the potential energy is smooth, then the assumption on the geometric ergodicity of GLA is typically satisfied for sufficiently small time-step.*

Given $\boldsymbol{Z}_k \in \mathbb{R}^{2n}$ and h, let $\vartheta_h : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ denote the exact time-h flow of Hamilton's equations (3). The *Exact Splitting* is defined as

$$\boldsymbol{Z}_{k+1} := \vartheta_h \circ \psi_{t_k+h, t_k}(\boldsymbol{Z}_k) \tag{8}$$

for $k = 0, ..., N-1$ with $\boldsymbol{Z}_0 = \boldsymbol{x}$.

Properties of GLA   The assumptions that appear in the following theorems are provided in §3.

Let $\mathbb{E}^{\boldsymbol{x}}\{\cdot\}$ denote the expectation conditioned on the initial condition being $\boldsymbol{x} \in \mathbb{R}^{2n}$. In terms of this notation, we can quantify the strong convergence of GLA to solution trajectories of inertial Langevin (1). The precise statement follows

Theorem 2.1 (Pathwise Accuracy). *Assume 3.1 and 3.2. For any $T > 0$, there exist $h_c > 0$ and $C(T) > 0$, such that for all $h < h_c$, $\boldsymbol{x} \in \mathbb{R}^{2n}$, and $t \in [0, T]$, GLA satisfies*

$$(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{X}_{\lfloor t/h \rfloor} - \boldsymbol{Y}(\lfloor t/h \rfloor h)|^2\})^{1/2} \leq C(T)(1 + |\boldsymbol{x}|^2)^{1/2} h. \qquad (9)$$

This result is expected because a Lie-Trotter splitting is first-order for deterministic ODEs, and the noise in (1) is additive.

Using this pathwise convergence, it is shown that GLA is geometrically ergodic with respect to a discrete invariant measure μ_h.

Theorem 2.2 (Geometric Ergodicity). *Assume 3.1, 3.2, and 3.3. Then GLA is geometrically ergodic with respect to a discrete invariant measure μ_h and the continuous Lyapunov function (cf. Assumption 3.3). That is, there exist $h_c > 0$, $\lambda > 0$, and $C_3 > 0$, such that for all $h < h_c$ and for all $k \geq 2$,*

$$|\mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_k)\} - \mu_h(f)| \leq C_3 V(\boldsymbol{x}) e^{-\lambda k h}, \quad \forall\ \boldsymbol{x} \in \mathbb{R}^{2n},$$

and for all test functions satisfying $|f(\boldsymbol{y})| \leq C_3 V(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}^{2n}$.

We stress this result is a consequence of strong convergence of GLA and the assumptions made on the potential energy and variational integrator. These assumptions are sufficient, but not necessary to guarantee this result.

Using geometric ergodicity we can quantify the equilibrium statistical accuracy of GLA. If p represents the global accuracy of GLA's underlying variational integrator, then μ_h is in TV distance $O(h^p)$ away from the Boltzmann-Gibbs measure μ. To be precise, the main result of the paper states

Theorem 2.3 (Long-Run Accuracy). *Assume 3.1, 3.2, and 3.3. Let μ_h denote the discrete invariant measure of GLA. Then, there exist $C > 0$ and $h_c > 0$, such that for all $h < h_c$,*

$$|\mu - \mu_h|_{TV} \leq C h^p.$$

There is a stronger argument in [23] based on the Feynman-Kac formula that can extend Theorem 2.3 to

$$|\mu(f) - \mu_h(f)| \leq C h^p, \qquad (10)$$

for all test functions $f \in L^2_\mu(\mathbb{R}^{2n})$ that are smooth with polynomial growth at infinity. The paper proves Theorem 2.3 with a more direct strategy. An

important point is that the proof is transparent since it involves a forward error analysis and does not rely on knowing the precise form of μ_h. The proof relies on the existence of μ_h and the nature of the convergence of GLA from a nonequilibrium position. Indeed, a backward error analysis of this discretization of the SDE (1) to characterize this invariant measure would be substantially more involved.

Implications   As a consequence of the TV error estimate derived in this paper, one can control the order of accuracy of μ_h by controlling the order of accuracy of GLA's underlying variational integrator. This is the distinguishing feature of GLA. Existing theory would indicate the accuracy of μ_h is the same order as the weak or strong accuracy of GLA. Theorem 2.1 states GLA is just first-order accurate on solution trajectories. Hence, existing theory would suggest that the equilibrium statistical accuracy of GLA is first-order, rather than pth-order accurate (where p is the order of accuracy of GLA's underlying variational integrator).

Existing theory would indicate to obtain a higher-order approximation of the invariant measure one would require a higher-order approximant to SDE (1) which entails approximation of multiple n-dimensional stochastic integrals per time-step. It is well-known that such higher-order discretizations of SDEs are computationally intensive. In contrast, a step of GLA requires evaluation of a single, n-dimensional stochastic integral per time-step. According to the main result of this paper, the order of accuracy of the variational integrator can be used to tune the TV-distance in Theorem 2.3 to a desired tolerance.

3   Preliminaries

The following assumptions on the potential energy, $U : Q \to \mathbb{R}$, will be used in this paper. These hypotheses are the same as those made in §7 of [10].

Assumption 3.1 (Assumptions on Potential Energy). *The potential energy function $U \in C^{\infty}(\mathbb{R}^n, \mathbb{R})$ satisfies:*

U1) *there exists a real constant $A_0 > 0$ such that*

$$|\nabla U(\boldsymbol{q}_0) - \nabla U(\boldsymbol{q}_1)| \leq A_0 |\boldsymbol{q}_0 - \boldsymbol{q}_1|, \ \ \forall \ \boldsymbol{q}_0, \boldsymbol{q}_1 \in \mathbb{R}^n.$$

8

U2) *there exists a real constant $A_1 > 0$ such that*

$$U(\boldsymbol{q}) \geq A_1(1 + |\boldsymbol{q}|^2), \quad \forall \ \boldsymbol{q} \in \mathbb{R}^n.$$

By standard results in stochastic analysis, condition **U1** is sufficient to guarantee almost sure existence and pathwise uniqueness of a solution to (1). The condition **U2** ensures that $e^{-\beta H}$ is integrable over $\mathbb{R}^{2n}$, and hence, that the Boltzmann-Gibbs measure is a well-defined probability measure. Assuming the solution to (1) is geometrically ergodic, we will prove in this paper that conditions **U1** - **U2** together with the following assumptions on the variational integrator, $\theta_h : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, are sufficient (but not necessary) to guarantee geometric ergodicty of GLA.

Assumption 3.2 (Assumptions on Variational Integrator). *For any $t > 0$ let ϑ_t denote the exact Hamiltonian flow of (3). The variational integrator $\theta_h : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ satisfies the following.*

V1) *θ_h is the discrete Hamiltonian map of a hyperregular discrete Lagrangian $L_d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ (cf. (52) and [9]).*

V2) *there exist constants $B_0 > 0$ and $h_c > 0$, such that for any $h < h_c$,*

$$|\theta_h(\boldsymbol{x}) - \vartheta_h(\boldsymbol{x})| \leq B_0(1 + |\boldsymbol{x}|^2)^{1/2}h^{p+1}, \quad \forall \ \boldsymbol{x} \in \mathbb{R}^{2n}.$$

As discussed in Appendix I, the condition **V1** implies that θ_h is symplectic, and hence, Lebesgue measure preserving. It will also be an important ingredient in proving Theorem 2.2 on geometric ergodicity of GLA. The condition **V2** states that the integrator is locally $(p+1)th$-order accurate.

Finally, we make the following structural assumption on (1).

Assumption 3.3 (Existence of a Lyapunov Function). *There exists $V \in C^\infty(\mathbb{R}^{2n}, \mathbb{R})$ and constants $C_i > 0$ such that*

$$C_0(1 + |\boldsymbol{x}|^2) \leq V(\boldsymbol{x}) \leq C_1(1 + |\boldsymbol{x}|^2), \quad \nabla V(\boldsymbol{x}) \leq C_2(1 + |\boldsymbol{x}|), \quad \forall \ \boldsymbol{x} \in \mathbb{R}^{2n},$$

$\lim_{\boldsymbol{x} \to \infty} V(\boldsymbol{x}) = \infty$, $a > 0$ and $c > 0$, such that for all $t > 0$,

$$\mathbb{E}^{\boldsymbol{x}}\{V(\boldsymbol{Y}(t))\} \leq e^{-at}V(\boldsymbol{x}) + \frac{c}{a}(1 - e^{-at}), \quad \forall \ \boldsymbol{x} \in \mathbb{R}^{2n}.$$

9

4   Analysis of GLA

4.1   Pathwise Convergence

Here GLA is shown to be first-order mean-squared convergent, which is a notion of pathwise convergence to solutions of (1) [15, 22]. The first-order accuracy of GLA on solution trajectories is not surprising because the method is derived from a Lie-Trotter splitting of (1). It is simply a generalization of the well-known fact that Lie-Trotter splittings of deterministic ODEs yield first-order accurate methods. This generalization is possible despite the lack of regularity in solutions because the noise in (1) is additive. Since the proof is standard, it will be kept terse.

Theorem 2.1 (Pathwise Accuracy). *Assume 3.1 and 3.2. For any $T > 0$, there exist $h_c > 0$ and $C(T) > 0$, such that for all $h < h_c$, $\boldsymbol{x} \in \mathbb{R}^{2n}$, and $t \in [0, T]$,*

$$(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{X}_{\lfloor t/h\rfloor} - \boldsymbol{Y}(\lfloor t/h\rfloor h)|^2\})^{1/2} \leq C(T)(1 + |\boldsymbol{x}|^2)^{1/2}h. \qquad (11)$$

Proof. By standard results in stochastic analysis, condition **U1** guarantees there a.s. exists a pathwise unique solution to (1): $\boldsymbol{Y}(t) \in \mathbb{R}^{2n}$ for $t \in [0, T]$ with $\boldsymbol{Y}(0) = \boldsymbol{x}$. Moreover, one can obtain the following bound on the second moment of of the solution: for all $T > 0$, there exists a $C(T) > 0$ such that for all $t \in [0, T]$,

$$\mathbb{E}^{\boldsymbol{x}}\left\{|\boldsymbol{Y}(t)|^2\right\} \leq C(T)(1 + |\boldsymbol{x}|^2). \qquad (12)$$

We will use this bound to invoke Theorem 1.1 in [15] which enables one to deduce global mean-squared error estimates of a discretization from local mean-squared error and local mean deviation. First, we establish this estimate for the exact splitting (8). By using Assumption **U1**, it is straightforward to show (see Lemma 7.1) that there exists $C > 0$ such that:

$$|\mathbb{E}^{\boldsymbol{x}}\{\boldsymbol{Y}(h) - \boldsymbol{Z}_1\}| \leq C \left(1 + |\boldsymbol{x}|^2\right)^{1/2} h^2 \qquad (13)$$

and

$$\left(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{Y}(h) - \boldsymbol{Z}_1|^2\}\right)^{1/2} \leq C \left(1 + |\boldsymbol{x}|^2\right)^{1/2} h^{3/2} \qquad (14)$$

Together with (12) this implies there exist $h_c > 0$ and $C(T) > 0$, such that for all $h < h_c$, $t \in [0, T]$ and $\boldsymbol{x} \in \mathbb{R}^{2n}$:

$$\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{Z}_{\lfloor t/h\rfloor}|^2\} \leq C(T)(1 + |\boldsymbol{x}|^2) \qquad (15)$$

10

Hence, by Theorem 1.1 in [15], one can show that for all $T > 0$, there exist $h_c > 0$ and $C(T) > 0$, such that for all $h < h_c$, $t \in [0, T]$ and $\boldsymbol{x} \in \mathbb{R}^{2n}$:

$$(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{Z}_{\lfloor t/h \rfloor} - \boldsymbol{Y}(\lfloor t/h \rfloor h)|^2\})^{1/2} \leq C(T) \left(1 + |\boldsymbol{x}|^2\right)^{1/2} h. \qquad (16)$$

Observe that the difference between a single step of GLA (7) and the exact splitting (8) can be written as

$$\boldsymbol{X}_1 - \boldsymbol{Z}_1 = (\theta_h - \vartheta_h) \circ \psi_{h,0}(\boldsymbol{x}).$$

Using Assumption **V2** one can show there exists $C > 0$ such that

$$\left(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{X}_1 - \boldsymbol{Z}_1|^2\}\right)^{1/2} \leq C \left(1 + |\boldsymbol{x}|^2\right)^{1/2} h^{p+1}, \qquad (17)$$

and, by Jensen's inequality:

$$|\mathbb{E}^{\boldsymbol{x}}\{\boldsymbol{X}_1 - \boldsymbol{Z}_1\}| \leq C \left(1 + |\boldsymbol{x}|^2\right)^{1/2} h^{p+1}. \qquad (18)$$

Together with (15) this implies that there exist $h_c > 0$ and $C(T) > 0$, such that for all $h < h_c$, $t \in [0, T]$ and $\boldsymbol{x} \in \mathbb{R}^{2n}$:

$$\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{X}_{\lfloor t/h \rfloor}|^2\} \leq C(T)(1 + |\boldsymbol{x}|^2) \qquad (19)$$

Using Assumption **U1** and Theorem 1.1 of [15], one can also show that for all $T > 0$, there exist $h_c > 0$ and $C(T) > 0$, such that for all $h < h_c$, $t \in [0, T]$ and $\boldsymbol{x} \in \mathbb{R}^{2n}$:

$$(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{X}_{\lfloor t/h \rfloor} - \boldsymbol{Z}_{\lfloor t/h \rfloor}|^2\})^{1/2} \leq C(T) \left(1 + |\boldsymbol{x}|^2\right)^{1/2} h^p. \qquad (20)$$

In other words, GLA is $O(h^p)$ strongly convergent to the exact splitting. One can then use the triangle inequality to obtain the estimate in the theorem from (20) and (16), i.e.,

$$(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{X}_{\lfloor t/h \rfloor} - \boldsymbol{Y}(\lfloor t/h \rfloor h)|^2\})^{1/2} \leq$$
$$\underbrace{(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{X}_{\lfloor t/h \rfloor} - \boldsymbol{Z}_{\lfloor t/h \rfloor}|^2\})^{1/2}}_{\leq K(T)(1+|\boldsymbol{x}|^2)^{1/2}h^p} + \underbrace{(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{Z}_{\lfloor t/h \rfloor} - \boldsymbol{Y}(\lfloor t/h \rfloor h)|^2\})^{1/2}}_{\leq K(T)(1+|\boldsymbol{x}|^2)^{1/2}h}.$$

In sum, GLA is first-order strongly convergent to solutions of (1). $\qquad \square$

11

4.2 Geometric Ergodicity

Geometric ergodicity is a strong type of stochastic stability of a Markov chain [13]. In this section geometric ergodicity of GLA is established following the recipe provided in §7 of [10]. In the context of this paper, geometric ergodicity means,

Definition 4.1. *A Markov chain $\boldsymbol{X}_k$ is said to be geometrically ergodic if there exist probability measure μ_∞, $\rho < 1$, and $M \in C^\infty(\mathbb{R}^{2n}, \mathbb{R}^+)$, such that*

$$|\mathbb{E}^{\boldsymbol{x}} \{f(\boldsymbol{X}_k)\} - \mu_\infty(f)| \le M(\boldsymbol{x})\rho^k, \quad \forall \; \boldsymbol{x} \in \mathbb{R}^{2n}, \quad \forall \; k \in \mathbb{N}, \qquad (21)$$

and for all $f \in L^2_{\mu_\infty}(\mathbb{R}^{2n})$ satisfying $|f(\boldsymbol{y})| \le M(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}^{2n}$.

Under the hypotheses below, the Lyapunov function from Assumption 3.3 is inherited by GLA.

Theorem 2.2 (Geometric Ergodicity)**.** *Assume 3.1, 3.2, and 3.3. Then GLA is geometrically ergodic with respect to a discrete invariant measure μ_h and the continuous Lyapunov function (cf. Assumption 3.3). That is, there exist $h_c > 0$, $\lambda > 0$, and $C_3 > 0$, such that for all $h < h_c$ and for all $k \ge 2$,*

$$|\mathbb{E}^{\boldsymbol{x}} \{f(\boldsymbol{X}_k)\} - \mu_h(f)| \le C_3 V(\boldsymbol{x})e^{-\lambda k h}, \quad \forall \; \boldsymbol{x} \in \mathbb{R}^{2n},$$

and for all test functions satisfying $|f(\boldsymbol{y})| \le C_3 V(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}^{2n}$.

Proof. This proof is an application of Theorem 2.5 of [10]. To invoke this theorem, we will show that GLA inherits the Lyapunov function $V : \mathbb{R}^{2n} \to \mathbb{R}$ of the continuous solution (cf. Assumption 3.3) and satisfies a minorization condition when sampled every other step.

To prove that GLA inherits the Lyapunov function $V : \mathbb{R}^{2n} \to \mathbb{R}$ we use Theorem 7.2 of [10]. This theorem assumes that the Lyapunov function of the SDE is essentially quadratic which follows from Assumption 3.3, and that the discretization of the SDE satisfies Condition 7.1 of [10]. Condition 7.1 (i) is a consequence of a single-step mean-squared error estimate of GLA which can be derived from (14) and (17). Condition 7.1 (ii) is satisfied for the first and second moments of GLA due to the estimate (19). Hence all of the assumptions of Theorem 7.2 [10] are satisfied, and one can conclude that GLA inherits the Lyapunov function $V : \mathbb{R}^{2n} \to \mathbb{R}$ up to a constant pre-factor.

Next, we prove that GLA satisfies a minorization condition when sampled every other step. This property follows from Lemma 2.3 of [10], because GLA sampled every other step admits a strictly positive, smooth transition probability function. In fact, this transition probability $q_h : \mathbb{R}^{2n} \times \mathbb{R}^{2n} \to [0,1]$ can be explicitly characterized, and by inspection it is clear that it is smooth as a function of its arguments and strictly positive everywhere.

To derive this expression, let $o_h : \mathbb{R}^n \times \mathbb{R}^n \to [0,1]$ denote the transition probability of the Ornstein-Uhlenbeck flow ψ_h (6). By a change of variables, it's transition density is given explicitly by:

$$o_h(\boldsymbol{p}_0, \boldsymbol{p}_1) =$$
$$\frac{1}{(2\pi)^{n/2} |\det(\boldsymbol{\Sigma}_h)|} \exp\left(-\frac{1}{2} \left(\boldsymbol{p}_1 - e^{-\gamma \boldsymbol{M}^{-1}h} \boldsymbol{p}_0 \right)^T \boldsymbol{\Sigma}_h^{-1} \left(\boldsymbol{p}_1 - e^{-\gamma \boldsymbol{M}^{-1}h} \boldsymbol{p}_0 \right) \right),$$
$$(22)$$

where
$$\boldsymbol{\Sigma}_h = \beta^{-1} \left(\boldsymbol{Id} - \exp(-2\gamma \boldsymbol{M}^{-1}h) \right) \boldsymbol{M}.$$

Let $\mathbf{D} = \mathbb{R}^{2n} \times \mathbb{R}^{2n} \times \mathbb{R}^{2n} \times \mathbb{R}^n$. Since the maps θ_h and ψ_h enjoy the Markov property, the transition probability of the composition $\theta_h \circ \psi_h \circ \theta_h \circ \psi_h$ can be expressed as a product of the transition probabilities of its components:

$$q_h\left((\boldsymbol{q}, \boldsymbol{p}), (\bar{\boldsymbol{q}}, \bar{\boldsymbol{p}})\right) =$$
$$\int_{\mathbf{D}} o_h(\boldsymbol{p}, \boldsymbol{p}_1)\delta((\boldsymbol{q}_1, \boldsymbol{p}_2) - \theta_h(\boldsymbol{q}, \boldsymbol{p}_1))o_h(\boldsymbol{p}_2, \boldsymbol{p}_3)\delta((\bar{\boldsymbol{q}}, \bar{\boldsymbol{p}}) - \theta_h(\boldsymbol{q}_1, \boldsymbol{p}_3))d\boldsymbol{p}_1 d\boldsymbol{p}_2 d\boldsymbol{p}_3 d\boldsymbol{q}_1.$$

The zero of the argument of the second Dirac-delta measure (from left) occurs at $(\boldsymbol{q}_1, \boldsymbol{p}_3) = \theta_h^{-1}(\bar{\boldsymbol{q}}, \bar{\boldsymbol{p}})$. Hence, the above expression simplifies,

$$q_h\left((\boldsymbol{q}, \boldsymbol{p}), (\bar{\boldsymbol{q}}, \bar{\boldsymbol{p}})\right) =$$
$$\int_{\mathbb{R}^{2n} \times \mathbb{R}^{2n}} o_h(\boldsymbol{p}, \boldsymbol{p}_1)\delta((\boldsymbol{q}_1, \boldsymbol{p}_2) - \theta_h(\boldsymbol{q}, \boldsymbol{p}_1))o_h(\boldsymbol{p}_2, \boldsymbol{p}_3)d\boldsymbol{p}_1 d\boldsymbol{p}_2.$$

By condition $\mathbf{V1}$ on θ_h, the zero of the argument of the remaining Dirac-delta measure above is uniquely determined by the discrete Hamiltonian flow of the discrete Lagrangian (cf. (51) in Appendix I). Hence, one obtains:

$$q_h\left((\boldsymbol{q}, \boldsymbol{p}), (\bar{\boldsymbol{q}}, \bar{\boldsymbol{p}})\right) =$$
$$|\det(D_{12}L_d(\boldsymbol{q}, \boldsymbol{q}_1, h))| o_h(\boldsymbol{p}, -D_1 L_d(\boldsymbol{q}, \boldsymbol{q}_1, h)) o_h(D_2 L_d(\boldsymbol{q}, \boldsymbol{q}_1, h), \boldsymbol{p}_3) \quad (23)$$

where $(\boldsymbol{q}_1, \boldsymbol{p}_3) = \theta_h^{-1}(\bar{\boldsymbol{q}}, \bar{\boldsymbol{p}})$. Using the hyperregularity assumption on the variational integrator **V1** (cf. (52)), (22), and (23), it is clear that q_h is a smooth probability transition function that is everywhere strictly positive. Hence, by Lemma 2.3 of [10], GLA sampled every other step satisfies a minorization condition.

In sum, we have shown that GLA satisfies a minorization condition and admits a Lyapunov function. The result follows from invoking Theorem 2.5 in [10]. $\qquad\square$

4.3 Long-Run Accuracy

Now we quantify the accuracy of GLA in sampling from the equilibrium measure of (1). For this purpose recall the following definition.

Definition 4.2 (Invariance of Measure). *A Markov chain $\boldsymbol{X}_k \in \mathbb{R}^{2n}$ is said to preserve a probability measure μ_∞ if for all $f \in L^2_{\mu_\infty}(\mathbb{R}^{2n})$ and $k \in \mathbb{N}$,*

$$\mathbb{E}_{\mu_\infty}\mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_k)\} = \mu_\infty(f) \tag{24}$$

where $\mu_\infty(f) = \int_{\mathbb{R}^{2n}} f d\mu_\infty$ and $\mathbb{E}_{\mu_\infty}\mathbb{E}^{\boldsymbol{x}}$ denotes expectation conditioned on the initial distribution being sampled from μ_∞, i.e.,

$$\mathbb{E}_{\mu_\infty}\mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_k)\} = \int_{\mathbb{R}^{2n}} \mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_k)\}\, \mu_\infty(d\boldsymbol{x}).$$

Given a step-size h, define the deviation GLA makes in preserving the Boltzmann-Gibbs measure, μ, as $\Delta_h^k : L^2_\mu(\mathbb{R}^{2n}) \to \mathbb{R}$:

$$\Delta_h^k(f) := \mathbb{E}_\mu\mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_k)\} - \mu(f).$$

Observe that if GLA exactly preserves μ then:

$$\Delta_h^k(f) = 0, \quad \forall\, f \in L^2_\mu(\mathbb{R}^{2n}).$$

The following local error result follows from the Ornstein-Uhlenbeck flow ψ_h preserving μ and the variational integrator θ_h preserving Lebesgue measure.

Lemma 4.3. *Suppose the potential energy satisfies* **U2**. *For a given $f \in L^2_\mu(\mathbb{R}^{2n})$,*

$$\Delta_h^1(f) = \int_{\mathbb{R}^{2n}} f(\boldsymbol{x}) \left(e^{-\beta\left(H((\theta_h)^{-1}(\boldsymbol{x})) - H(\boldsymbol{x}) \right)} - 1 \right) \mu(d\boldsymbol{x}).$$

14

Proof. The condition **U2** ensures that μ is a well-defined probability measure. According to the definition of GLA (7), $\boldsymbol{X}_1 = \theta_h \circ \psi_h(\boldsymbol{x})$. Substitute this expression into Δ_h^1 to obtain:

$$\Delta_h^1(f) = \int_{\mathbb{R}^{2n}} \mathbb{E}^{\boldsymbol{x}} \{f(\theta_h \circ \psi_h(\boldsymbol{x}))\} \mu(d\boldsymbol{x}) - \int_{\mathbb{R}^{2n}} f d\mu.$$

Since ψ_h preserves μ and θ_h is deterministic it follows that,

$$\Delta_h^1(f) = \int_{\mathbb{R}^{2n}} f(\theta_h(\boldsymbol{x}))\mu(d\boldsymbol{x}) - \int_{\mathbb{R}^{2n}} f d\mu.$$

Changing variables under the map θ_h in the first integral above, and using the volume-preserving property of the variational integrator θ_h (See Appendix.) one obtains the desired expression. $\square$

Remark 4.1. *As a consequence of Lemma 4.3, if θ_h admits no energy error, then GLA preserves μ. In particular, the exact splitting (8) preserves μ.*

In the situation where GLA is geometrically ergodic, this paragraph quantifies the equilibrium error of GLA in preserving the BG measure.

Lemma 4.4. *Assume 3.1, 3.2, and 3.3. Then, there exist $C > 0$ and $h_c > 0$, such that for all $h < h_c$,*

$$\lim_{N \to \infty} \left| \Delta_h^N(f) \right| \leq Ch^p,$$

and for all $f \in L_{\mu_\infty}^2(\mathbb{R}^{2n})$ satisfying $|f(\boldsymbol{y})| \leq C_3 V(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}^{2n}$.

Proof. Let $f \in L_\mu^2(\mathbb{R}^{2n})$ such that $|f(\boldsymbol{y})| \leq C_3 V(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}^{2n}$. The term $\mathbb{E}_\mu \mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_N)\}$ can be written as a telescoping sum:

$$\mathbb{E}_\mu \mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_N)\} = \mu(f) + \sum_{k=1}^{N} \left(\mathbb{E}_\mu \mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_k)\} - \mathbb{E}_\mu \mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_{k-1})\} \right).$$

By Lemma 4.3, one can rewrite/reindex this sum as:

$$\Delta_h^N(f) = \int_{\mathbb{R}^{2n}} \sum_{k=0}^{N-1} \mathbb{E}^{\boldsymbol{x}} \{f(\boldsymbol{X}_k)\} \left(e^{-\beta(H(\theta_h^{-1}(\boldsymbol{x})) - H(\boldsymbol{x}))} - 1 \right) \mu(d\boldsymbol{x}). \qquad (25)$$

15

Since θ_h preserves Lebesgue measure, one can write this deviation as:

$$\Delta_h^N(f) =$$
$$\int_{\mathbb{R}^{2n}} \sum_{k=0}^{N-1} \underbrace{(\mathbb{E}^{\boldsymbol{x}}\left\{f(\boldsymbol{X}_k)\right\} - \mu_h(f))}_{\text{Deviation from Equilibrium}} \cdot \underbrace{\left(e^{-\beta(H(\theta_h^{-1}(\boldsymbol{x})) - H(\boldsymbol{x}))} - 1\right)}_{\text{Energy Error of Variational Integrator}} \mu(d\boldsymbol{x}).$$

$$(26)$$

From (26) it is clear that the equilibrium BG error is due to: 1) how fast GLA converges to equilibrium and 2) the local accuracy with which θ_h represents the Hamiltonian function H. The equality (26) is the crux of the proof, and what follows is an approach to bound $\Delta_h^N(f)$.

Since GLA is geometrically ergodic (cf. Theorem 2.2), one can bound $\Delta_h^N(f)$ from above by

$$\left|\Delta_h^N(f)\right| \leq \left(\sum_{k=0}^{N-1} e^{-\lambda h k}\right) C_3 \int_{\mathbb{R}^{2n}} V(\boldsymbol{x}) \left|e^{-\beta(H(\theta_h^{-1}(\boldsymbol{x})) - H(\boldsymbol{x}))} - 1\right| \mu(d\boldsymbol{x}).$$

Changing variables in the right-hand-side under the map θ_h, one can rewrite this bound as,

$$\left|\Delta_h^N(f)\right| \leq \left(\sum_{k=0}^{N-1} e^{-\lambda h k}\right) C_3 \int_{\mathbb{R}^{2n}} V(\theta_h(\boldsymbol{x})) \left|e^{-\beta(H(\theta_h(\boldsymbol{x})) - H(\boldsymbol{x}))} - 1\right| \mu(d\boldsymbol{x}).$$

In the limit as $N \to \infty$, the right-hand-side of the above can be written in terms of the formula for the geometric series for $e^{-\lambda h}$:

$$\lim_{N\to\infty} \left|\Delta_h^N(f)\right| \leq \frac{C_3}{1 - e^{-\lambda h}} \int_{\mathbb{R}^{2n}} V(\theta_h(\boldsymbol{x})) \left|e^{-\beta(H(\theta_h(\boldsymbol{x})) - H(\boldsymbol{x}))} - 1\right| \mu(d\boldsymbol{x}). \quad (27)$$

Using the natural bound $|e^x - 1| \leq e^{|x|} - 1$ for all $x \in \mathbb{R}$, one can further bound $|\Delta_h^N(f)|$ by:

$$\lim_{N\to\infty} \left|\Delta_h^N(f)\right| \leq \frac{C_3}{1 - e^{-\lambda h}} \int_{\mathbb{R}^{2n}} V(\theta_h(\boldsymbol{x})) \left(e^{\beta|H(\theta_h(\boldsymbol{x})) - H(\boldsymbol{x})|} - 1\right) \mu(d\boldsymbol{x}). \quad (28)$$

Introduce the exact flow ϑ_h of Hamilton's equations (3) into this bound,

$$\lim_{N\to\infty} \left|\Delta_h^N(f)\right| \leq \frac{C_3}{1 - e^{-\lambda h}} \int_{\mathbb{R}^{2n}} V(\theta_h(\boldsymbol{x})) \left(e^{\beta|H(\theta_h(\boldsymbol{x})) - H(\vartheta_h(\boldsymbol{x}))|} - 1\right) \mu(d\boldsymbol{x}).$$

$$(29)$$

Set $\boldsymbol{y}_0 = \theta_h(\boldsymbol{x})$ and $\boldsymbol{y}_1 = \vartheta_h(\boldsymbol{x})$. By the fundamental theorem of calculus,

$$H(\boldsymbol{y}_1) - H(\boldsymbol{y}_0) = \int_0^1 \nabla H(\boldsymbol{y}_0 + s(\boldsymbol{y}_1 - \boldsymbol{y}_0)) \cdot (\boldsymbol{y}_1 - \boldsymbol{y}_0) ds.$$

Using condition **U1** and the Cauchy-Schwartz inequality, it follows from the above that there exists $C > 0$ such that

$$|H(\boldsymbol{y}_1) - H(\boldsymbol{y}_0)| \leq C(1 + |\boldsymbol{y}_1| + |\boldsymbol{y}_0|)|\boldsymbol{y}_1 - \boldsymbol{y}_0|.$$

Another application of the condition **U1** and **V2** implies there exists $C > 0$ such that

$$|H(\boldsymbol{y}_1) - H(\boldsymbol{y}_0)| \leq C(1 + |\boldsymbol{x}|^2)h^{p+1}.$$

Therefore,

$$\lim_{N \to \infty} |\Delta_h^N(f)| \leq \frac{C_3}{1 - e^{-\lambda h}} \int_{\mathbb{R}^{2n}} V(\theta_h(\boldsymbol{x})) \left(e^{\beta K(1+|\boldsymbol{x}|^2)h^{p+1}} - 1 \right) \mu(d\boldsymbol{x}). \quad (30)$$

Now we show how the the factor $V(\theta_h(\boldsymbol{x}))$ above is handled.

Since the Lyapunov function is quadratically bounded, the variational integrator satisfies **V2**, and the Hamiltonian vector field is uniformly Lipschitz by condition **U1**, there exists $C > 0$ such that

$$\lim_{N \to \infty} |\Delta_h^N(f)| \leq \frac{C}{1 - e^{-\lambda h}} \int_{\mathbb{R}^{2n}} (1 + |\boldsymbol{x}|^2) \left(e^{\beta K(1+|\boldsymbol{x}|^2)h^{p+1}} - 1 \right) \mu(d\boldsymbol{x}). \quad (31)$$

By condition **U2** the total energy is quadratically bounded from below. Consequently one can bound $e^{-\beta H(\boldsymbol{x})}$ by $e^{-\beta D(1+|\boldsymbol{x}|^2)}$ for some constant $D > 0$. Thus,

$$\lim_{N \to \infty} |\Delta_h^N(f)| \leq$$
$$\frac{C}{1 - e^{-\lambda h}} \int_{\mathbb{R}^{2n}} (1 + |\boldsymbol{x}|^2) \left(e^{\beta K(1+|\boldsymbol{x}|^2)h^{p+1}} - 1 \right) e^{-\beta D(1+|\boldsymbol{x}|^2)} d\boldsymbol{x}.$$

When $h < h_c = (D/K)^{1/(p+1)}$ the above integral is finite and one obtains the desired error estimate. $\qquad\square$

A simple application of Theorem 2.3 implies an error estimate for μ_h. For this purpose we introduce the total variation between measures μ and ν:

$$|\mu - \nu|_{TV} = \sup_{|f| \leq 1} \left| \int_{\mathbb{R}^{2n}} f(\boldsymbol{x})(\mu(d\boldsymbol{x}) - \nu(d\boldsymbol{x})) \right|.$$

17

Since $\tilde{M}(\boldsymbol{y}) \geq 1$ for all $\boldsymbol{y} \in \mathbb{R}^{2n}$, Theorem 2.3 applies for all $f \in L_\mu^2(\mathbb{R}^{2n})$ such that $|f(\boldsymbol{y})| \leq 1$ for all $\boldsymbol{y} \in \mathbb{R}^{2n}$. The TV norm can be written as:

$$|\mu - \mu_h|_{TV} =$$
$$\sup_{|f| \leq 1} \left| \int_{\mathbb{R}^{2n}} f d\mu - \mathbb{E}_\mu \mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_N)\} + \mathbb{E}_\mu \mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_N)\} - \int_{\mathbb{R}^{2n}} f d\mu_h \right|$$

By the triangle inequality,

$$|\mu - \mu_h|_{TV} \leq \sup_{|f| \leq 1} \left| \Delta_h^N(f) \right| + \sup_{|f| \leq 1} \left| \mathbb{E}_\mu \mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_N)\} - \mu_h(f) \right|. \qquad (32)$$

However, under the hypotheses of the theorem, GLA is geometrically ergodic with respect to μ_h and hence,

$$\lim_{N \to \infty} \sup_{|f| \leq 1} |\mathbb{E}_\mu \mathbb{E}^{\boldsymbol{x}}\{f(\boldsymbol{X}_N)\} - \mu_h(f)| \to 0 \qquad (33)$$

and,

$$|\mu - \mu_h|_{TV} \leq \lim_{N \to \infty} \sup_{|f| \leq 1} \left| \Delta_h^N(f) \right|. \qquad (34)$$

Lemma 4.4 can now be invoked to obtain from (34) an upper bound for the TV distance between μ and μ_h. This concludes the proof of Theorem 2.3 which we restate:

Theorem 2.3 (Long-Run Accuracy). *Assume 3.1, 3.2, and 3.3. Let μ_h denote the discrete invariant measure of GLA. Then, there exist $C > 0$ and $h_c > 0$, such that for all $h < h_c$,*

$$|\mu - \mu_h|_{TV} \leq Ch^p.$$

In summary, the preceding analysis showed the TV error estimate in Theorem 2.3 relies on GLA's variational integrator θ_h being volume-preserving and *pth*-order accurate, the Ornstein-Uhlenbeck map ψ_h exactly preserving the Boltzmann-Gibbs measure, and GLA being geometrically ergodic. To establish the latter, we used the strategy adopted in [10] which relates pathwise convergence of a discretization of an SDE to geometric ergodicity of the discretization. This strategy requires the potential force is uniformly Lipschitz.

5 Validation

This section tests three different instances of GLA on a variety of simple mechanical systems governed by Langevin equations. The purpose of this section is to confirm the error estimates provided in the paper.

Let h be a fixed step size and $\xi_k \sim \mathcal{N}(0,1)$ for $k \in \mathbb{N}$. The following update scheme is obtained by composing the explicit first-order, symplectic Euler method with ψ_h:

$$
\begin{cases}
\hat{p}_k & = & e^{-\gamma h} p_k + \sqrt{\frac{1-e^{-2\gamma h}}{\beta}} \xi_k, \\
q_{k+1} & = & q_k + h\hat{p}_k, \\
p_{k+1} & = & \hat{p}_k - h\frac{\partial U}{\partial q}(q_{k+1}),
\end{cases}
\tag{35}
$$

for $k \in \mathbb{N}$. The following integrator is obtained by composing the second-order accurate explicit, symmetric, symplectic Störmer-Verlet method with ψ_h:

$$
\begin{cases}
\hat{p}_k & = & e^{-\gamma h} p_k + \sqrt{\frac{1-e^{-2\gamma h}}{\beta}} \xi_k, \\
P_k^{1/2} & = & \hat{p}_k - \frac{h}{2}\frac{\partial U}{\partial q}(q_k), \\
q_{k+1} & = & q_k + hP_k^{1/2}, \\
p_{k+1} & = & P_k^{1/2} - \frac{h}{2}\frac{\partial U}{\partial q}(q_{k+1}),
\end{cases}
\tag{36}
$$

for $k \in \mathbb{N}$. The following integrator is obtained by composing a fourth-order accurate explicit, symmetric, symplectic method due to F. Neri (see, e.g., [27]) with ψ_h:

$$
\begin{cases}
Q_1 & = & q_k, \\
P_1 & = & e^{-\gamma h} p_k + \sqrt{\frac{1-e^{-2\gamma h}}{\beta}} \xi_k, \\
\begin{cases}
P_{i+1} & = & P_i - c_i h\frac{\partial U}{\partial q}(Q_i), \\
Q_{i+1} & = & Q_i + d_i h P_{i+1},
\end{cases} & \quad i = 1, ..., 4, \\
q_{k+1} & = & Q_5, \\
p_{k+1} & = & P_5,
\end{cases}
\tag{37}
$$

for $k \in \mathbb{N}$, and where we have introduced the following constants:

$$
c_1 = c_4 = \frac{1}{2(2-2^{1/3})}, \quad c_2 = c_3 = \frac{1-2^{1/3}}{2(2-2^{1/3})},
$$

$$
d_1 = d_3 = \frac{1}{2-2^{1/3}}, \quad d_2 = \frac{-2^{1/3}}{2-2^{1/3}}, d_4 = 0.
$$

The purpose of this fourth-order symplectic integrator is for validation. For "optimal" fourth and fifth-order accurate symplectic integrators that minimize the error in the Hamiltonian, the reader is referred to [11].

We will show that despite the fact that (36) and (37) are only first-order pathwise convergent according to Theorem 2.1, they approximate ensemble averages of μ-integrable functions that satisfy $|f(q,p)| \leq M(q,p)$ for all $(q,p) \in \mathbb{R}^{2n}$ to within second and fourth-order accuracy, respectively. This is consistent with Theorem 2.3.

Linear Oscillator   This section follows the analysis of numerical methods for linear oscillators governed by Langevin equations developed in [6,15]. The governing equations for a linear oscillator of unit mass at uniform temperature $1/\beta$ are given explicitly by evaluating (1) at $U(q) = q^2/2$:

$$\begin{cases} dq &= pdt, \\ dp &= -qdt - \gamma pdt + \sqrt{2\beta^{-1}\gamma}dW. \end{cases} \tag{38}$$

The resulting process is Gaussian with stationary distribution given by the BG distribution:

$$P_\infty(q,p) = Z^{-1}\exp\left(-\beta\left(\frac{p^2}{2} + \frac{q^2}{2}\right)\right)$$

and with

$$\mu(q^2) = \lim_{t\to\infty}\mathbb{E}\{q_t^2\} = 1/\beta, \quad \mu(p^2) = \lim_{t\to\infty}\mathbb{E}\{p_t^2\} = 1/\beta, \quad \kappa(qp) = \lim_{t\to\infty}\mathbb{E}\{q_t p_t\} = 0.$$

The stationary distribution of the geometric Langevin integrators (35)-(37) is also Gaussian with equilibrium distribution of the form:

$$P_h(q,p) = \frac{1}{2\pi|\Sigma^{-1}|}\exp\left(-\frac{1}{2}\begin{pmatrix} q & p \end{pmatrix}\Sigma^{-1}\begin{pmatrix} q \\ p \end{pmatrix}\right)$$

where

$$\Sigma = \begin{bmatrix} \sigma_q^2 & \kappa \\ \kappa & \sigma_p^2 \end{bmatrix}, \quad \sigma_q^2 = \lim_{n\to\infty}\mathbb{E}\{q_n^2\}, \quad \sigma_p^2 = \lim_{n\to\infty}\mathbb{E}\{p_n^2\}, \quad \kappa = \lim_{n\to\infty}\mathbb{E}\{q_n p_n\}.$$

This stationary correlation matrix can be explicitly determined. For (35) its entries are given by:

$$\sigma_q^2 = \frac{\left(1 + e^{\gamma h}\right)^2}{\left(2 + 2e^{\gamma h} - h^2\right)\beta} = \frac{1}{\beta} + \mathcal{O}(h)$$

$$\sigma_p^2 = \frac{2 + 2e^{\gamma h} - h^2 + e^{2\gamma h}h^2}{\left(2 + 2e^{\gamma h} - h^2\right)\beta} = \frac{1}{\beta} + \mathcal{O}(h^2)$$

$$\gamma = -\frac{e^{\gamma h}\left(1 + e^{\gamma h}\right)h}{\left(2 + 2e^{\gamma h} - h^2\right)\beta} = \mathcal{O}(h)$$

Observe that the cumulative error (35) makes is of $\mathcal{O}(h)$, i.e.,

$$|\sigma_q^2 - \mu(q^2)| + |\sigma_p^2 - \mu(p)^2)| + |\kappa - \mu(qp)| \leq \mathcal{O}(h).$$

Whereas for (36) its entries are given by:

$$\sigma_q^2 = \frac{4}{\beta(4 - h^2)} = \frac{1}{\beta} + \frac{h^2}{4\beta} + \mathcal{O}(h^4)$$

$$\sigma_p^2 = \frac{1}{\beta}$$

$$\kappa = 0$$

and its cumulative error is of $\mathcal{O}(h^2)$, i.e.,

$$|\sigma_q^2 - \mu(q^2)| + |\sigma_p^2 - \mu(p)^2)| + |\kappa - \mu(qp)| \leq \mathcal{O}(h^2).$$

For (37) its entries are given by:

$$\sigma_q^2 = \frac{1}{\beta} + \frac{\left(-4 - 3 \times \sqrt[3]{2} - 2 \times 2^{2/3}\right)h^4}{144\beta} + O(h^5)$$

$$\sigma_p^2 = \frac{1}{\beta}$$

$$\kappa = 0$$

and its cumulative error is of $\mathcal{O}(h^4)$, i.e.,

$$|\sigma_q^2 - \mu(q^2)| + |\sigma_p^2 - \mu(p)^2)| + |\kappa - \mu(qp)| \leq \mathcal{O}(h^4).$$

Finally, consider the exact splitting applied to the linear oscillator at uniform temperature. Hamilton's equations for a linear oscillator are:

$$\begin{bmatrix} \dot{q} \\ \dot{p} \end{bmatrix}(t) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix}(t), \quad \begin{bmatrix} q \\ p \end{bmatrix}(0) = \begin{bmatrix} q_0 \\ p_0 \end{bmatrix},$$

with explicit solution given by:

$$\begin{bmatrix} q \\ p \end{bmatrix}(t) = \begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} q_0 \\ p_0 \end{bmatrix}.$$

Thus, the exact splitting update is given by:

$$\begin{bmatrix} q_1 \\ p_1 \end{bmatrix} = \begin{bmatrix} \cos(h) & \sin(h) \\ -\sin(h) & \cos(h) \end{bmatrix} \begin{bmatrix} q_0 \\ p_0 \end{bmatrix} + \sqrt{\frac{e^{2\gamma h} - 1}{\beta}} \begin{bmatrix} \sin(h) \\ \cos(h) \end{bmatrix} \xi_0.$$

In this situation one can show there is no error made in the stationary correlation matrix. This follows from the fact that the exact solution of Hamilton's equations is volume and energy preserving.

Nonglobally Lipschitz, Nonlinear Oscillator The theory in this paper does not apply to this example since the potential force is nonglobally Lipschitz. With a nonglobally Lipschitz potential force, for any $h > 0$ there will exist regions in phase space where the Lipschitz constant of the potential force is beyond the linear stability threshold of an explicit variational integrator θ_h. Hence, a GLA based on an explicit variational integrator will be stochastically unstable; transient, to be precise. However, for the step-sizes and variational integrators employed, and for the duration of the numerical experiments, discrete orbits of GLA seem to be confined to a compact region of phase space where the variational integrator θ_h is linearly stable and Monte Carlo estimates are consistent with the error estimates in the paper.

The governing equations for a cubic oscillator of unit mass at uniform temperature $1/\beta$ are given explicitly by evaluating (1) at $U(q) = q^4/4 - q^2/2$:

$$\begin{cases} dq &= pdt, \\ dp &= (q - q^3)dt - \gamma pdt + \sqrt{2\beta^{-1}\gamma}dW. \end{cases} \tag{39}$$

The resulting potential force is only locally Lipschitz.

22

Time-Step	Number of Steps	(35)	(36)	(37)
h	N	3.11e-02	8.03e-03	1.45e-02
h/2	2 N	1.49e-02	1.94e-03	9.80e-04
h/4	4 N	7.42e-03	4.83e-04	7.35e-05
h/8	8 N	3.74e-03	1.29e-04	5.79e-06

Table 1: The table estimates $|\mu_h(q^2) - \mu(q^2)|$ using empirical time-averages with $N = 40 \times 10^9$ steps and $h = 0.4$ with GLA as determined by (35)-(37). For subsequent rows the time-steps are halved and the number of steps doubled, so that the time-interval of integration is fixed for all experiments. The results show that as the time-steps are halved the difference decreases linearly for (35), nearly quadratically for (36), and nearly quartically for (37). These results are consistent with the error estimates in the paper.

The estimates shown earlier predict that

$$|\mu(q^2) - \mu_h(q^2)| \leq \mathcal{O}(h^p)$$

where p is the order of accuracy of θ_h. Hence, one expects near fourth-order accuracy for (37), near second-order accuracy for (36) and first-order accuracy for (35) as shown in table 1. The tests will apply (35)-(37) to estimate

$$\lim_{t \to \infty} \mathbb{E}\{q_t^2\} = \mu(q^2) = \frac{\int_{-\infty}^{\infty} q^2 e^{-\beta U(q)} dq}{\int_{-\infty}^{\infty} e^{-\beta U(q)} dq}$$

by empirical averages of the form

$$I^{h,N} := \frac{1}{N}\left(\sum_{i=1}^{N} q_i^2\right).$$

As nicely discussed in [23], in addition to the discretization error $|\mu(q^2) - \mu_h(q^2)|$ one has to cope with the statistical error arising from the time-average being finite, i.e., $I^{h,N} \approx \mu_h(q^2)$. The computations were performed with $\gamma = 1$ and an inverse temperature value of $\beta = 2$.

6 Conclusion

The analysis in this paper represents a first step towards a deeper analysis of GLA for molecular systems. In this paper we make assumptions on the

Hamiltonian that ensure the solution to inertial Langevin and GLA are geometrically ergodic. In particular, we assume the Hamiltonian vector field is uniformly Lipschitz and the Hamiltonian is coercive. These hypotheses are sufficient to ensure GLA is geometrically ergodic whenever the solution process is. In particular, the former hypothesis is important to ensure GLA is stochastically stable [13]. If GLA's underlying variational integrator is not globally linearly stable, one can show GLA defines a transient Markov chain. Still one can use GLA as proposal step within a Metropolis-Hastings algorithm to obtain a stochastically stable Metropolis-Adjusted Geometric Langevin Algorithm (MAGLA). A numerical analysis of MAGLA including pathwise convergence can be found in [4].

A closer inspection of the proof of Theorem 2.3 reveals that the estimate relies on the following important ingredients:

1. GLA is geometrically ergodic with respect to a probability measure μ_h;

2. the variational integrator is Lebesgue-measure preserving;

3. the Ornstein-Uhlenbeck flow preserves μ; and,

4. the local energy error of the variational integrator is $(p+1)th$-order accurate.

Therefore, we stress that the result holds under more general conditions. The main point being:

> *If GLA is geometrically ergodic with respect to a unique invariant measure, the error in sampling the invariant measure of the SDE is determined by the energy error in GLA's variational integrator.*

7 Appendix

7.1 Single-Step Error

Lemma 7.1. *Assume 3.1 and 3.2. For h small enough, there exists a $C > 0$ such that*

$$|\mathbb{E}^{\boldsymbol{x}}\{\boldsymbol{Y}(h) - \boldsymbol{Z}_1\}| \leq C \left(1 + |\boldsymbol{x}|^2\right)^{1/2} h^2 \tag{40}$$

and

$$\left(\mathbb{E}^{\boldsymbol{x}}\{|\boldsymbol{Y}(h) - \boldsymbol{Z}_1|^2\}\right)^{1/2} \leq C \left(1 + |\boldsymbol{x}|^2\right)^{1/2} h^{3/2}. \tag{41}$$

Proof. Write $\mathbf{Y}(t) = (\boldsymbol{Q}(t), \boldsymbol{P}(t))$ where $\boldsymbol{Q}(t)$ and $\boldsymbol{P}(t)$ represent the instantaneous configuration and momentum of the system, respectively. In terms of which write the SDE (1) as:

$$\begin{cases} d\boldsymbol{Q}/dt &= \boldsymbol{M}^{-1}\boldsymbol{P} \\ d\boldsymbol{P} &= -\nabla U(\boldsymbol{Q})dt - \gamma \boldsymbol{M}^{-1}\boldsymbol{P}dt + \sqrt{2\gamma\beta^{-1}}d\boldsymbol{W} \end{cases} \tag{42}$$

$\boldsymbol{Q}(0) = \boldsymbol{Q}_0$ and $\boldsymbol{P}(0) = \boldsymbol{P}_0$. It will be useful to write out the solution of (42). For this purpose integrate (42) to obtain:

$$\boldsymbol{Q}(h) = \boldsymbol{Q}_0 + h\boldsymbol{M}^{-1}\boldsymbol{P}_0 + \int_0^h \boldsymbol{M}^{-1}[-\nabla U(\boldsymbol{Q}(s)) - \gamma \boldsymbol{M}^{-1}\boldsymbol{P}(s)](h - s)ds$$

$$+ \sqrt{2\gamma\beta^{-1}} \int_0^h (h - s)\boldsymbol{M}^{-1}d\boldsymbol{W}(s) \tag{43}$$

and

$$\boldsymbol{P}(h) = e^{-\gamma \boldsymbol{M}^{-1}h}\boldsymbol{P}_0 - h\nabla U(\boldsymbol{Q}_0) - \int_0^h (h - s)\frac{\partial^2 U}{\partial \boldsymbol{q}^2}(\boldsymbol{Q}(s)) \cdot \boldsymbol{M}^{-1}\boldsymbol{P}(s)ds$$

$$+ \int_0^h (\boldsymbol{I} - e^{-\gamma \boldsymbol{M}^{-1}(h-s)})\nabla U(\boldsymbol{Q}(s))ds + \boldsymbol{\eta} \tag{44}$$

where we have introduced:

$$\boldsymbol{\eta} = \sqrt{2\gamma\beta^{-1}} \int_0^h e^{-\gamma \boldsymbol{M}^{-1}(h-s)}d\boldsymbol{W}(s).$$

Write $\mathbf{Z}(t) = (\hat{\boldsymbol{Q}}(t), \hat{\boldsymbol{P}}(t))$ where $\hat{\boldsymbol{Q}}(t)$ and $\hat{\boldsymbol{P}}(t)$ represent the instantaneous configuration and momentum of the exact splitting, respectively. The exact splitting after a single step solves

$$\begin{cases} d\hat{\boldsymbol{Q}}/dt &= \boldsymbol{M}^{-1}\hat{\boldsymbol{P}} \\ d\hat{\boldsymbol{P}}/dt &= -\nabla U(\hat{\boldsymbol{Q}}) \end{cases} \tag{45}$$

where $\hat{\boldsymbol{Q}}(0) = \boldsymbol{Q}_0$ and $\hat{\boldsymbol{P}}(0) = e^{-\gamma \boldsymbol{M}^{-1}h}\boldsymbol{P}_0 + \boldsymbol{\eta}$. Integrating (45) yields,

$$\hat{\boldsymbol{Q}}(h) = \boldsymbol{Q}_0 + h\boldsymbol{M}^{-1}e^{-\gamma \boldsymbol{M}^{-1}h}\boldsymbol{P}_0 - \int_0^h \boldsymbol{M}^{-1}\nabla U(\hat{\boldsymbol{Q}}(s))(h - s)ds + h\boldsymbol{M}^{-1}\boldsymbol{\eta} \tag{46}$$

and

$$\hat{\boldsymbol{P}}(h) = e^{-\gamma \boldsymbol{M}^{-1} h} \boldsymbol{P}_0 - h \nabla U(\boldsymbol{Q}_0)$$
$$- \int_0^h (h - s) \frac{\partial^2 U}{\partial \boldsymbol{q}^2} (\hat{\boldsymbol{Q}}(s)) \cdot \boldsymbol{M}^{-1} \hat{\boldsymbol{P}}(s) ds + \boldsymbol{\eta}. \tag{47}$$

To obtain the mean-squared and mean error estimates we will use the following bounds on the second moment of the continuous solution and the exact splitting. Namely, for all $t \in [0, h]$, there exists a $C > 0$ such that

$$\mathbb{E}^{\boldsymbol{x}} \left\{ |\boldsymbol{Z}(t)|^2 \right\} \vee \mathbb{E}^{\boldsymbol{x}} \left\{ |\boldsymbol{Y}(t)|^2 \right\} \leq C(1 + |\boldsymbol{x}|^2) \tag{48}$$

where $\boldsymbol{x} = (\boldsymbol{Q}_0, \boldsymbol{P}_0)$. We will prove this estimate for the exact splitting, and omit the proof for the continuous solution since it is very similar. Let $\hat{\boldsymbol{x}} = (\hat{\boldsymbol{Q}}(0), \hat{\boldsymbol{P}}(0))$. By Taylor's formula,

$$|\boldsymbol{Z}(t)|^2 = |\hat{\boldsymbol{x}}|^2 + 2 \int_0^t \left\langle \hat{\boldsymbol{Q}}(s), \boldsymbol{M}^{-1} \hat{\boldsymbol{P}}(s) \right\rangle ds + 2 \int_0^t \left\langle \hat{\boldsymbol{P}}(s), -\nabla U(\hat{\boldsymbol{Q}}(s)) \right\rangle ds$$

By Young's inequality,

$$|\boldsymbol{Z}(t)|^2 \leq |\hat{\boldsymbol{x}}|^2$$
$$+ \int_0^t (|\hat{\boldsymbol{Q}}(s)|^2 + |\boldsymbol{M}^{-1} \hat{\boldsymbol{P}}(s)|^2) ds + \int_0^t (|\hat{\boldsymbol{P}}(s)|^2 + |\nabla U(\hat{\boldsymbol{Q}}(s))|^2) ds$$

The uniform Lipschitz condition **U1** implies a linear growth condition on the potential force. Hence, there exists a constant $C > 0$ such that

$$|\boldsymbol{Z}(t)|^2 \leq |\hat{\boldsymbol{x}}|^2 + C \int_0^t |\hat{\boldsymbol{Z}}(s)|^2 ds$$

By Gronwall's lemma it follows that,

$$|\boldsymbol{Z}(t)|^2 \leq |\hat{\boldsymbol{x}}|^2 e^{Ch}$$

for $t \leq h$. Hence, for h small enough we obtain the desired bound on the second moment of the exact splitting.

The difference between (43) and (46) is,

$$
\begin{aligned}
\boldsymbol{Q}(h) - \hat{\boldsymbol{Q}}(h) = \\
h\boldsymbol{M}^{-1}(\boldsymbol{I} - e^{-\gamma \boldsymbol{M}^{-1}h})\boldsymbol{P}_0 \\
+ \int_0^h \boldsymbol{M}^{-1}[\nabla U(\hat{\boldsymbol{Q}}(s)) - \nabla U(\boldsymbol{Q}(s))](h-s)ds \\
+ \int_0^h \boldsymbol{M}^{-1}[-\gamma \boldsymbol{P}(s)ds + \sqrt{2\gamma\beta^{-1}}d\boldsymbol{W}(s)](h-s) - h\boldsymbol{M}^{-1}\boldsymbol{\eta} \quad (49)
\end{aligned}
$$

Likewise, the difference between (44) and (47) is,

$$
\begin{aligned}
\boldsymbol{P}(h) - \hat{\boldsymbol{P}}(h) = \\
\int_0^h (h-s)\left[\frac{\partial^2 U}{\partial \boldsymbol{q}^2}(\hat{\boldsymbol{Q}}(s)) \cdot \boldsymbol{M}^{-1}\hat{\boldsymbol{P}}(s) - \frac{\partial^2 U}{\partial \boldsymbol{q}^2}(\boldsymbol{Q}(s)) \cdot \boldsymbol{M}^{-1}\boldsymbol{P}(s)\right]ds \\
+ \int_0^h (e^{-\gamma \boldsymbol{M}^{-1}(h-s)} - \boldsymbol{I})\nabla U(\boldsymbol{Q}(s))ds \quad (50)
\end{aligned}
$$

From (49) and (50), it is clear that the leading term of the expectation of these differences is $O(h^2)$ and the leading term in the mean-squared expectation of the differences is $O(h^{3/2})$. To bound these terms one needs the bounds on the second moments of the solutions and the exact splitting provided in (48). To enable estimation of (50) one needs control of the Hessian of U. The assumption of smoothness on U and the uniform Lipschitz condition **U1** on the potential force provide this control. In particular, since a differentiable function is Lipschitz continuous if and only if it has bounded differential, the Frobenius norm of the Hessian of U is bounded by the Lipschitz constant of the potential force.

$\square$

7.2 Variational Integrators

Let $L : \mathbb{R}^{2n} \to \mathbb{R}$ denote the Lagrangian obtained from the Legendre transform of the Hamiltonian H, and given by:

$$
L(\boldsymbol{q}, \boldsymbol{v}) = \frac{1}{2}\boldsymbol{v}^T \boldsymbol{M}\boldsymbol{v} - U(\boldsymbol{q}).
$$

A variational integrator is defined by a discrete Lagrangian $L_d : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}$ which is an approximation to the so-called *exact discrete Lagrangian*

which is defined as:

$$L_d^E(\boldsymbol{q}_0, \boldsymbol{q}_1, h) = \int_0^h L(\boldsymbol{Q}, \dot{\boldsymbol{Q}}) dt$$

where $\boldsymbol{Q}(t)$ solves the Euler-Lagrange equations for the Lagrangian L with endpoint conditions $\boldsymbol{Q}(0) = \boldsymbol{q}_0$ and $\boldsymbol{Q}(h) = \boldsymbol{q}_1$.

By passing to the Hamiltonian description, a discrete Lagrangian determines a symplectic integrator on $\mathbb{R}^{2n}$ as follows. Given $(\boldsymbol{q}_0, \boldsymbol{p}_0) \in \mathbb{R}^{2n}$, a variational integrator defines an update $(\boldsymbol{q}_1, \boldsymbol{p}_1) \in \mathbb{R}^{2n}$ by the following system of equations:

$$\begin{cases} \boldsymbol{p}_0 & = -D_1 L_d(\boldsymbol{q}_0, \boldsymbol{q}_1, h), \\ \boldsymbol{p}_1 & = D_2 L_d(\boldsymbol{q}_0, \boldsymbol{q}_1, h). \end{cases} \tag{51}$$

Denote this map by $\theta_h : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, i.e.,

$$\theta_h : \quad (\boldsymbol{q}_0, \boldsymbol{p}_0) \mapsto (\boldsymbol{q}_1, \boldsymbol{p}_1),$$

where $(\boldsymbol{q}_1, \boldsymbol{p}_1)$ solve (51). One can show that θ_h preserves the canonical symplectic form on $\mathbb{R}^{2n}$, and hence, is Lebesgue measure preserving [9]. By appropriately constructing L_d, the map θ_h can define an approximation to the flow of Hamilton's equations for the Hamiltonian H (3). Hyperregularity of the discrete Lagrangian means for all $h > 0$

$$|\det D_{12} L_d(\boldsymbol{q}_0, \boldsymbol{q}_1, h)| > 0, \quad \forall \, \boldsymbol{q}_0, \boldsymbol{q}_1 \in Q \times Q. \tag{52}$$

References

[1] E. Akhmatskaya, N. Bou-Rabee, and S. Reich, *A comparison of generalized hybrid Monte Carlo methods with and without momentum flip*, J. Comp. Phys. **228** (2009), 2256–2265.

[2] N. Bou-Rabee and H. Owhadi, *Boltzmann-Gibbs preserving Langevin integrators*, arXiv:0712.4123v3 [math.NA], 2009.

[3] _____, *Stochastic variational integrators*, IMA J. of Numer. Anal. **29** (2009), 421–443.

[4] N. Bou-Rabee and E. Vanden-Eijnden, *Pathwise accuracy of Metropolized integrators for SDEs*, CPAM (2009), arXiv:0905.4218v1 [math.NA].

[5] A. Brünger, C. L. Brooks, and M. Karplus, *Stochastic boundary conditions for molecular dynamics simulations of ST2 water*, Chem. Phys. Lett. **105** (1984), 495–500.

[6] K. Burrage, I. Lenane, and G. Lythe, *Numerical methods for second-order stochastic differential equations*, SIAM J. Sci. Comp. **29** (2007), 245–264.

[7] G. Bussi and M. Parrinello, *Accurate sampling using Langevin dynamics*, Phys. Rev. E **75** (2007), 056707.

[8] A. Lew, J. E. Marsden, M. Ortiz, and M. West, *Asynchronous variational integrators*, Arch. Ration. Mech. An. **167** (2003), 85–145.

[9] J. E. Marsden and M. West, *Discrete mechanics and variational integrators*, Acta Numerica **10** (2001), 357–514.

[10] J. C. Mattingly, A. M. Stuart, and D. J. Higham, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stoch. Proc. Appl. **101** (2002), no. 2, 185–232.

[11] R. I. McLachlan and P. Atela, *The accuracy of symplectic integrators*, Nonlinearity **4** (1992), 541–562.

[12] R. I. McLachlan and M. Perlmutter, *Conformal Hamiltonian systems*, Journal of Geometry and Physics **39** (2001), 276–300.

[13] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, 2nd ed., Cambridge University Press, New York, NY, 2009.

[14] G. N. Milstein and M. V. Tretyakov, *Quasi-symplectic methods for Langevin-type equations*, IMA J. Num. Anal. **23** (2003), 593–626.

[15] _____, *Stochastic numerics for mathematical physics*, Springer, Berlin, 2004.

[16] G. A. Pavliotis, A. M. Stuart, and K. C. Zygalakis, *Calculating effective diffusivities in the limit of vanishing molecular diffusion*, J. Comput. Phys. **228** (2008), 1030–1055.

[17] A. Scemama, T. Lelièvre, G. Stoltz, E. Cancés, and M. Caffarel, *An efficient sampling algorithm for variational Monte Carlo*, J. Chem. Phys. **125** (2006), 114105.

[18] M. Serrano, G. De Fabritiis, P. Espanol, and P. V. Coveney, *A stochastic Trotter integration scheme for dissipative particle dynamics*, Math. Comput. Simulat. **72** (2006), 190–194.

[19] T. Shardlow, *Splitting for dissipative particle dynamics*, SIAM J. Sci. Comput. **24** (2003), 12671282.

[20] R. D. Skeel and J. Izaguirre, *An impulse integrator for Langevin dynamics*, Mol. Phys. **100** (2002), 3885–3891.

[21] C. Soize, *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady-state solutions*, Series on Advances in Mathematics for Applied Sciences, 17, World Scientific, Singapore, 1994.

[22] D. Talay, *Simulation and numerical analysis of stochastic differential systems : a review*, Probabilistic Methods in Applied Physics (P. Krèe and W. Wedig, eds.), vol. 451, Springer-Verlag, Berlin, 1995, pp. 54–96.

[23] _____, *Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme*, Markov Processes and Related Fields **8** (2002), 1–36.

[24] W. F. van Gunsteren and H. J. C. Berendsen, *Algorithms for brownian dynamics*, Mol. Phys. **45** (1982), 637–647.

[25] E. Vanden-Eijnden and G. Ciccotti, *Second-order integrators for Langevin equations with holonomic constraints*, Chem. Phys. Lett. **429** (2006), 310–316.

[26] J. M. Wendlandt and J. E. Marsden, *Mechanical integrators derived from a discrete variational principle*, Physica D **106** (1997), 223–246.

[27] H. Yoshida, *Construction of higher-order symplectic integrators*, Phys. Lett. A **150** (1990), 262–268.